

9 Digital Libraries and the Social Web: Scholarship

Karen Calhoun

Cornell University Library (retired)

ksc10@cornell.edu

Note: This is a preprint of a chapter whose final and definitive form was co-published in *Exploring Digital Libraries: Foundations, Practice, Prospects* by [Facet Publishing](#) (2014) and [ALA Neal-Schuman](#) (2014).

Overview

This chapter is the first of two that consider the responses of digital libraries to the social web and to web-based practices for information seeking, learning, teaching, research, professional recognition, work, recreation and socializing. Both chapters elaborate on ideas introduced in chapters 3 to 8. This chapter's first part examines the origins and chaotic, fast-moving nature of the social web, explores the possibility of digital libraries as social platforms, and introduces a visual framework that attempts to bring some coherence to the many confusing elements of digital libraries' evolution toward the social web. The second part of the chapter turns to the branches of the visual framework that pertain to the social web's existing, emergent, or potential impacts on scholarship, research and researchers.

Introduction

As discussed in chapter 6, the public library framework developed by McClure (1987) casts library roles in terms of direct involvement in the communities to be served. Their framework is liberating in that it shifts attention away from an information-processing or collection-centric definition of libraries toward a community-centric definition. In the same way, the advent of the social web provides an opportunity to shift the focus and core assumptions of digital libraries away from their collections and information processes (selecting, organizing, providing access, etc.) in favor of new ways of thinking about services, expectations and potential social roles.

Keywords: Digital libraries—Social aspects; Online social networks; Scholarly communication; Bibliometrics; Author identification; Researcher profiling; Reputation

Before considering specific opportunities afforded by the social web, it is necessary to spend some time considering its background.

Background: Web 2.0 and Library 2.0

Tim O'Reilly dates the first use of the phrase "web 2.0" to a conversation he had with Dale Dougherty at a 2004 technology conference (O'Reilly 2005). They intended the phrase to mark not just a turning point for the web following the dot-com collapse of 2001, but also to capture new notions of ***the web as a platform for participation***, characterized by:

- machine-to-machine interactions (syndication, web services and mashups)
- significant user interface innovation
- a bias for two-way "conversation" rather than one-way information dissemination
- radically democratized and collaborative content creation and exchange
- open, reusable data

While the phrase web 2.0 became a buzzword for many, the concept captured enough meaning to fire both imaginations and actions that have transformed the web from a "flat, read-only kaleidoscope" (web 1.0) to a "shared collaboratory" (Hammond, Hannay and Lund 2004, characterizing Tim Berners-Lee's original conception of the web as a collaboration space).

Michael Casey, a librarian at Gwinnet County Public Library in Georgia, was one of those inspired by the concept of web 2.0 and its implications for libraries. He launched the first "Library 2.0" blog, called LibraryCrunch, in September 2005 (Casey and Savastinuk 2007, xxi). Others quickly picked up and further developed the concept (among them Miller 2005; Maness 2006; Stephens 2006; Habib 2006; Blyberg 2006; Levine 2006; Abram 2007). A professional debate emerged almost as quickly about the meaning and scope of Library 2.0. Did Library 2.0 simply refer to web-based technologies applied to library services and collections (e.g., tagging in the

online catalog)? Or, as suggested by Blyberg (2006), was Library 2.0 inherently disruptive, fundamentally challenging the library assumptions and service models of the time? The debate was not resolved, and as of this writing, the use of the phrase “Library 2.0” in the literature appears to be diminishing (a Google Scholar search for articles with “library 2.0” in the title suggests the possibility that use of the phrase may have peaked in the period 2007-2008). Nevertheless as the societal impact of web 2.0 has grown to massive scale, many individuals have embraced the practices and values of this “platform for participation”—and they bring their expectations with them when they visit libraries, including digital libraries.

What is the social web?

For the purpose of this book, the term “social web” refers collectively to the web sites, tools and services that facilitate interactions, collaboration, content creation and sharing, contribution and participation on the web. The distinguishing characteristic is human interaction: the social web supports many types of online communities, and not just those who participate in social networks. Their tools include e-mail, listservs, bookmarking, wikis and blogs, microblogs, photo or video sharing services, e-meeting rooms, review sites and more. Vickery and Wunsch-Vincent’s extensive report for the OECD (2007), worldwide in scope, documents and explains the participative web and its tools supporting user-created content at that time. Despite its age, the report remains informative for its definitions, data and multinational perspective.

It is worth interjecting a few words about wikis here, since the concept for them dates to 1994 (“History of wikis” 2013) and they are so significant to the emergence of the social web and the high value it places on user-contributed content. More than any other feature of the social web perhaps, wikis exemplify the global shift to using the web as a platform for participation. Wikipedia (launched 2001) is of course the best-known example of a wiki. The definition of a

“wiki” is “a website that allows users to add, remove or otherwise edit and change content collectively” (Vickery and Wunsch-Vincent 2007, 33, 37).

In addition to the applications that are visible to users, a number of underlying machine-to-machine tools create the foundation for social web interactions, for example web services, APIs and mashups (discussed in chapter 4). These allow servers in different places to exchange or combine services or content. This means that something that “lives” in one place on the web (like a video, a calendar, or Twitter comments) can be dynamically shared, posted and updated in many places.

In addition to the many web services and APIs that support the social web, the large-scale take-up of mobile smartphones, tablets and other mobile devices has created a huge scope of opportunity for social web growth. The market for mobile application development (mobile apps) is large. As of this writing the latest Pew Internet Project reports indicate that 56% of American adults owned a smartphone; 34% owned a tablet; half reported having apps on their phones and 82% had them on tablets (Anderson and Rainie 2012; pewinternet.org, trends, device ownership, May 2013). In 2011 researchers began reporting that Americans spend more time using mobile apps than they do browsing the web using their mobile devices (Walsh 2011). Chapter 10 returns to the discussion of mobile apps in the context of digital libraries.

Digital libraries and the social web

Chakraborty, in his overview of digital libraries and the social web, points out that despite the rise of personalized, interactive online environments, most digital libraries continue to operate from a traditional, collections-centered service model (2010, 127). Brusilovsky and others (2010, 116) make the point that “the social nature of the library is typically lost when the library goes digital.” Indeed, the first 15 years or so of digital library work produced mostly read-only

(“web 1.0”) digital libraries, and a digital library that incorporates social software applications continues to be the exception rather than the rule. The typical digital library’s service model aligns with the conventional, collections-centered library worldview discussed in chapter 4.

Yet digital library users now expect more than rich collections. Hull, Pettifer and Kell (2008) describe scholarly digital libraries as “cold, impersonal and isolated” as well as poorly integrated for human and machine interaction. They contrast such “frozen” digital libraries with more social, interactive tools for scholars like Zotero, Mendeley, CiteULike and others, in which content moves fluidly between web applications. They make the case that such tools better support typical researcher workflows and methods for collaboration.

The social digital library?

Digital libraries are now faced with finding their place in the fast-moving, chaotic information space that is the social web. So far, digital libraries have not been displaced. However at this point in digital library evolution, it has become a pressing matter to not only adopt but embrace the social web’s “principles of participation,” as advocated by Lankes and colleagues for libraries in general (2007, 31). A number of digital library experts have also persistently called for aligning digital libraries with social web principles and practices. Notably, Candela and others (2007, 6) of the DELOS Network of Excellence on Digital Libraries published its *Digital Library Manifesto*, which redefined the notion of a digital library as (emphasis added):

...a ***tool at the center of intellectual activity*** having no logical, conceptual, physical, temporal, or personal borders or barriers to information. Generally accepted conceptions have ***shifted from a content-centric system*** that merely supports the organization and provision of access to particular collections of data and information, ***to a person-centric system...***

Even before the DELOs Manifesto appeared, some digital library experts had begun to explore the feasibility and utility of web 2.0 protocols and more participatory frameworks for digital libraries. Here it is worth mentioning the example of the Ockham Initiative (Morgan, Frumpkin and Fox 2004), an early digital library project that explored the possibilities of web-based registries, web services and social tools (e.g., annotating and reviewing) in a digital library environment. There are many other examples of early work: mining the digital library literature from 2003 forward produces a fair number of articles exploring social web concepts, attributes and tools. Subsequent sections of this chapter discuss a few of these articles.

Digital libraries as social platforms

Dan Cohen (2010), in a recorded lecture at Cambridge University on the social life of digital libraries, urged a transition from static repositories to social platforms that are active, open, modular, gregarious and “chatty” with other software, servers, people and organizations. A number of digital library technologists, for example those working on the Hydra projects (projecthydra.org) discussed in chapter 4, are exploring more open repository frameworks that could perhaps help with such a transition.

The nature of digital libraries is not incompatible with new roles as social systems. As discussed in chapter 1, the Kahn-Wilensky architecture of digital libraries does not constrain them to being read-only repositories that support only search and retrieval services, nor does the architectural model exclude the incorporation of a variety of components and agents (see Altman 2008, 154). In his analysis of digital libraries, Altman predicted the expansion of digital libraries’ personalization and collaborative roles. There is nothing in the intrinsic nature of digital libraries blocking the way forward.

What may make the transition of digital libraries to social platforms difficult is the weight of libraries' long-successful traditions, core values and practices. One of the digital library experts interviewed for this book noted "the values of the social web (for example, 'everyone is a creator of content') are hard for librarians to integrate with their own values of authority and authenticity." It is not that processes for vetting the credibility of sources are no longer respected or important; however these approaches sit (at times uncomfortably) alongside the predominant values of the social web: engagement and participation.

The evolution of digital libraries' responses to the social web can be understood as a set of continuums with distinct extremes but many variations along the transition (figure 9.1). Prior sections have already discussed the continuums from the focus on collections to the focus on networked communities, from repositories to social platforms, and from the values of authority and authentication to those of engagement and participation. Figure 9.1 proposes several others.

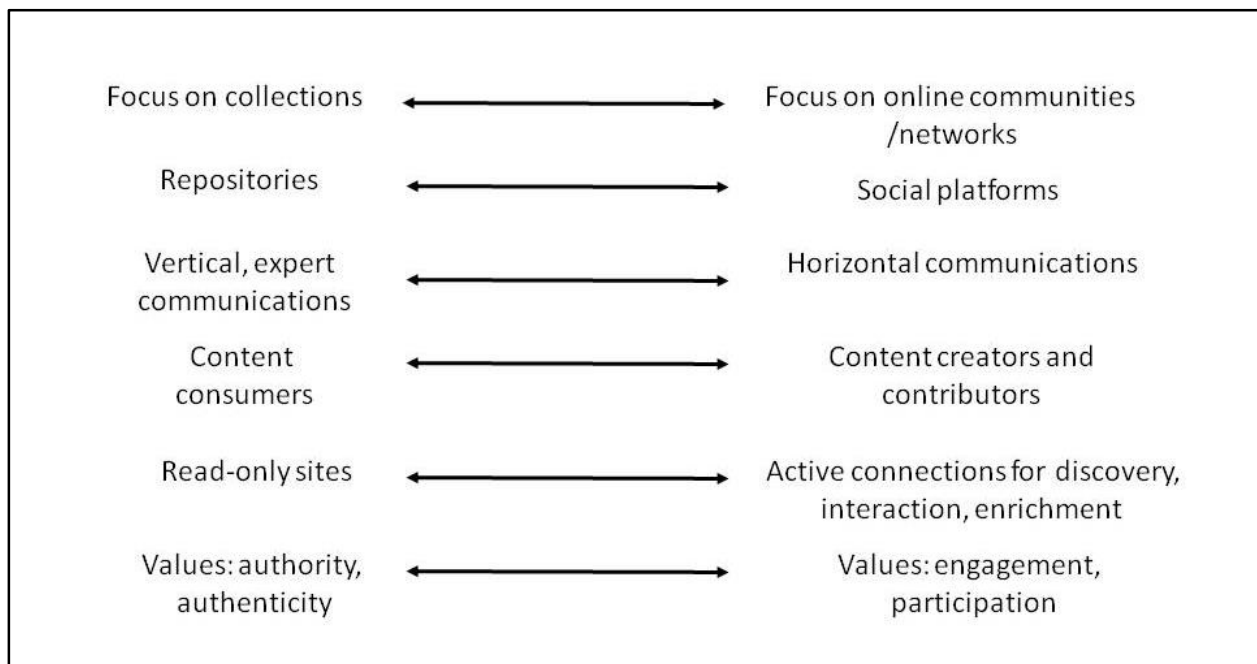


Figure 9.1 Transitions associated with the shift to social digital libraries

Librarianship is not the only profession affected by the shift to the social web's spirit of collaboration and personal self-efficacy; the impact on medicine and health care—where the use of relevant, credible sources is so critical—has perhaps been stronger. Gunther Eysenbach, a leading researcher in e-health, the internet and medicine, discusses the powerful impact of the social web on the medical and health care fields in his frequently cited article on “Medicine 2.0” (Eysenbach 2008). He describes how the social web has provided alternatives for patients to discover relevant, trustworthy information through a disintermediated process that taps into the collective wisdom of the social web, taking the form of shared bookmarks, recommender systems, wikis, social networks and other web tools.

Digital libraries' social evolution: a visual framework

Figure 9.2 is an attempt to visualize key drivers and components of the evolution of digital libraries toward new roles on the social web as the branches and roots of a tree. The left side of the map suggests the present situation for most digital libraries and other distinguishing characteristics of the current landscape. The right side of the map represents opportunities (some already realized) for a stronger social web presence and more robust community roles for digital libraries. The map places ideas related to the evolution of the scholarship as the upper branches of the tree. The middle branches represent innovations and opportunities for digital library collections to become more visible to and interactive with online communities. The roots of the tree illustrate the shifts in the core values and assumptions of digital libraries.

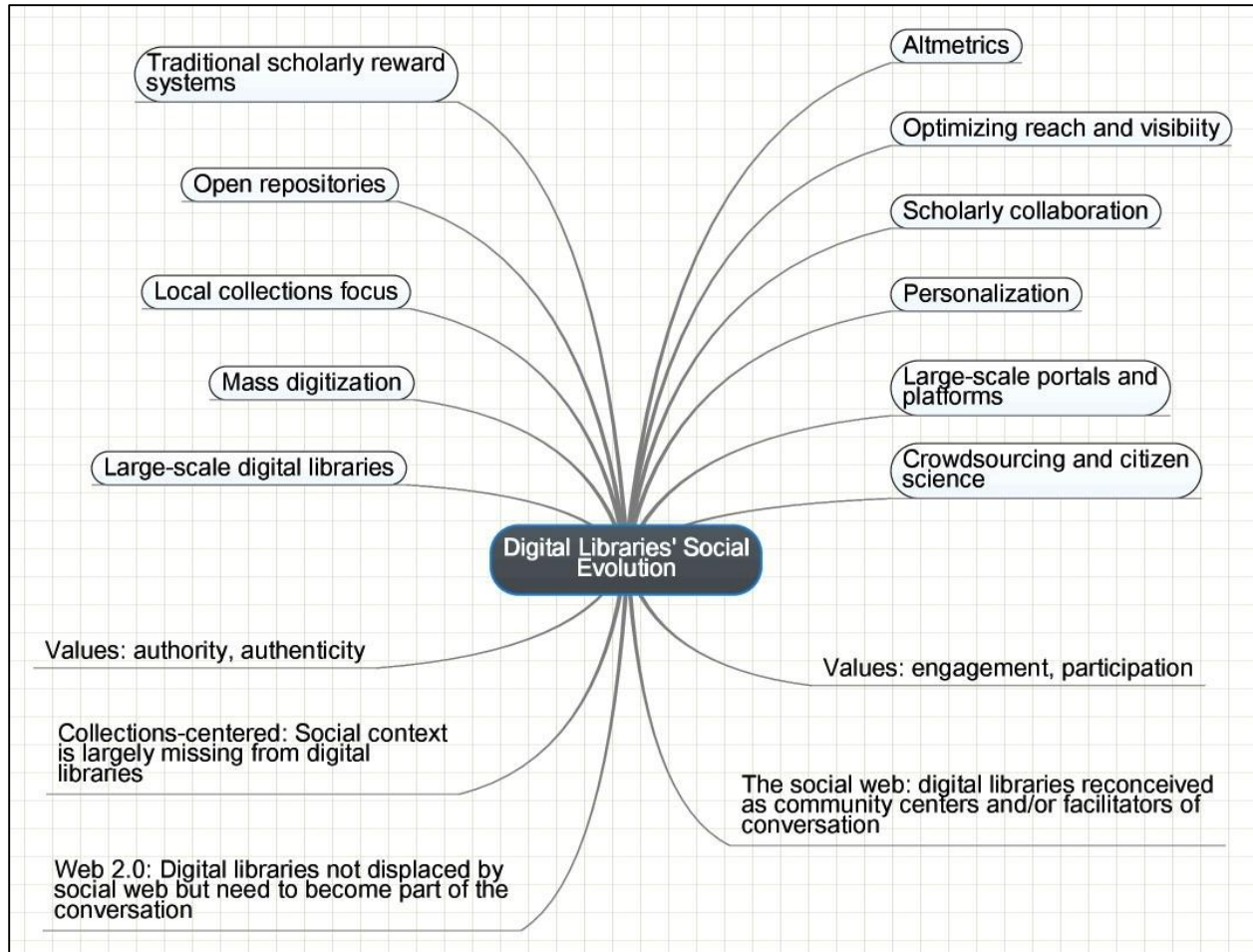


Figure 9.2 The evolution of digital libraries toward new roles on the social web

Scholarship

The second part of this chapter further explores the branches of figure 9.2 that pertain to the social web's existing, emergent, or potential impacts on researchers and scholarship. Figure 9.3 visualizes a few examples, categories or opportunities associated with each of these branches. Individual sections then discuss each branch. The left side of figure 9.3 visualizes the current situation; the right side visualizes innovations, experiments or possibilities.

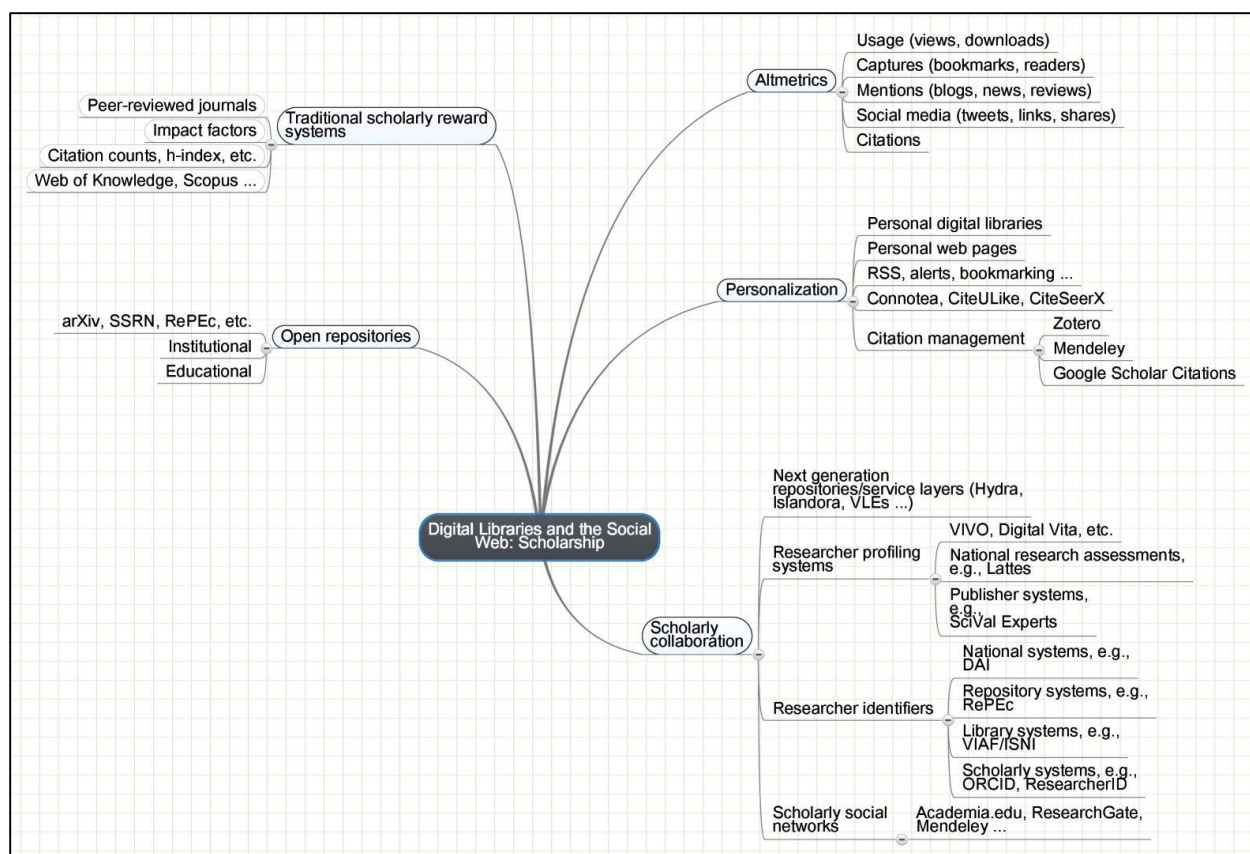


Figure 9.3 Social web impacts on researchers and scholarship

The traditional scholarly reward system

It is obvious that the processes of scholarly communication have changed radically. But how much have scholars changed? Thinking back to the definition of “scholarly communication” in

chapter 1, and aside from the radically-changed *processes* they use, how much change has there been in scholars' values and preferences as writers, linkers, information disseminators and collaborators? This is not so clear. Oya Rieger's socio-cultural analyses of institutional repositories (2007; 2008a) suggest that the attitudes and preferences of faculty, researchers and graduate students are shaped by centuries-old scholarly practices and norms. The roles of commercial publishers and peer review in certifying the quality of new scholarly assets are deeply entrenched, and they provide the foundation for a scholar's promotion and tenure. Journal impact factors and citation counts can be expected to be at the heart of scholarly reward systems for the foreseeable future.

Rieger emphasizes that trying to move faculty and "their deeply embedded value systems" directly to new technologies and forms of scholarly communication will fail. More recently, the latest Ithaka longitudinal study of US faculty research behavior and preferences reveals the persistent importance of peer-reviewed journals with high impact factors to scholars' decisions of where to publish their work (Schonfeld, Housewright and Wulfson 2013, 58-60). The study found little evidence for a decline in the importance of traditional, closed scholarly publishing practices in favor of more open and social dissemination of new research.

Chapter 8 discusses the fresh momentum for open access publication produced by the UK's Finch Report and the policies and targets announced by the European Commission in 2012. There may nevertheless be a gap between the UK and EU policy-driven emphasis on open access and the current perspectives of many UK and European scholars. As an example, a UK study using the same methodology as the US Ithaka study found that UK scholars, like US scholars, place high importance on reaching their academic peers using traditional methods involving peer review and high impact journals, and less importance on reaching the general

public and making their work freely available on the internet (Housewright, Schonfeld and Wulfson 2013, 8, 69-71).

At the same time, as in the fields of medicine and health care, the social web is providing valuable complements to traditional closed, hierarchical systems for pointing to trustworthy scholarly content. Like everyone else, scholars are participating in the social web and using its tools (such as blogs, shared bookmarking services, Twitter, Mendeley and more). As discussed in chapter 4, open access repositories and search engine indexing are already disclosing an unprecedented amount of scholarly content to the public. It is becoming possible to take advantage of the collective wisdom represented by choices made using social web tools. In addition it is becoming possible to evaluate the credentials of scholars and researchers in new ways, for example using specialized social networks and researcher profiling systems (such as LinkedIn, Academia.edu, ResearchGate.net, Mendeley, Google Scholar Citations, VIVO, SciVal Experts and more). Scholarly outputs are also assessed using systems developed to support national research assessment exercises such as those discussed later in this chapter.

Altmetrics

A new subfield of bibliometrics (quantitative analysis of scholarly communication) is emerging, called “altmetrics” (new quantitative measures based on the social web; see Roemer and Borchardt 2012 for an introduction to altmetrics). Jason Priem, who as of this writing is an information science PhD student, coined the term in a tweet in 2010 (Priem, Piwowar and Hemminger 2012, reference 4). The intention is to use altmetrics alongside formal, acknowledged indicators of influence like citation counts. Priem and others’ 2012 paper in arXiv.org reports their early results indicating the viability of altmetrics as a method for capturing the impact of papers. They studied social media tools and citation counts as sources of data related to over 24,000 articles in PLoS journals (their table 1 lists the data sources, which

included two shared bookmarking sites, Twitter, Wikipedia, Facebook, several blogs, downloads, citations, comments and rankings). They conclude that there is sufficient social web data to fuel the altmetrics approach; that with citation data, altmetrics can more fully describe scholarly impact; and that altmetrics can provide insight into different types of impact on different audiences.

As of this writing there is a great deal of conversation and activity focused on altmetrics; however it is early days to determine where the field will go from here. The value of altmetrics is being debated; for example Judy Luther's blog post on the subject (2012), which captured the main points of the controversy, attracted 40 comments, some forcefully stated. Richard Cave's presentation at a recent Charleston Conference (2012, slide 8) provides the typology of altmetrics data sources used in figure 9.3.

Open repositories

While they are not social sites, the most successful subject-based repositories (arXiv, RePEc, SSRN, etc.) may be viewed as harbingers of the emergent online scholarly communities of the social web. As discussed in chapter 8, the most successful subject-based repositories have grown organically around the scholarly communities they serve (see the examples in chapters 2, 4, 6 and 7), and they are woven into the way their disciplines communicate. These repositories have evolved through the online community life cycle (described in chapter 7), from a strong community orientation at inception to a high degree of trust and participation at maturity.

The progress of early educational digital libraries (for example, some of those spawned by the NSDL project, discussed in chapter 7) illustrated some painful aspects of these repositories' social evolution, as some of their builders struggled to engage communities of teachers and

learners. The evidence presented in multiple chapters makes it clear that institutional repositories continue to struggle to engage their intended communities. Well-funded social web alternatives are beginning to emerge and attract participation by researchers (e.g., scholarly social networks). These developments increase the pressure on institutional repositories to find firmer footing among the services supporting the creation and dissemination of scholarship. Their most promising strategies (illustrated on the right side of figure 9.3) appear to be evolution toward (1) next-generation repository platforms, as described in chapter 4, and (2) better reach and visibility on high-traffic sites, including sites where the content can be not only discovered and consumed but also reused, annotated, bookmarked and shared.

From personalization to collaboration

Personal digital libraries

Neil Beagrie, a digital preservation expert from the UK, had the foresight to recognize the immense impact that the trend toward individual digital creativity, coupled with the availability of large digital storage and computing power to individuals, would have on where digital content comes from, who collects it, and how it is stored and shared (Beagrie 2005). Noting that “people are able to create, capture and store an ever-increasing amount of digital information about or for themselves,” Beagrie articulated early the connection between web-based personalization and collaboration services and platforms. He argued that personal digital collections (such as emails, collected documents, alerts and bookmarks, personal webpages, blogs, portfolios of work, digital images, audio and video recordings and more) would engender new services for easily marking, tagging, annotating, editing, sharing and/or distributing them on the web and thereby “reinforce informal social networks and mechanisms of communication.”

Tony Hammond and colleagues of the Nature Publishing Group recognized as early as 2004 the game-changing impact of content syndication tools like RSS (a web-based format for

automated, immediate sharing and distribution of web content, like headlines or blog posts, to RSS feed subscribers to that site). The Nature Publishing Group and other science publishers began taking up RSS as a means of alerting interested readers to new content; the updates could be directed not only to desktops but to mobile devices (Hammond, Hannay and Lund 2004).

RSS feeds and other innovations

RSS feeds joined a number of other innovations that had appeared by 2005 or 2006 to enable first, personalization and soon after, social web collaboration. Along with wikis, blogs and other tools, the means for social web collaboration included bookmarking, whose history predates search engines, as described by Hammond and others (2005). The advantages to be had from collaboratively sharing bookmarks quickly spurred the introduction of various social bookmarking services, like delicious.com (founded in 2003), as well as a number of tools with an academic focus, like Connotea (Lund et al. 2005; in 2013 Baynes announced Connotea's retirement) or CiteULike.org (developed in 2004 and still highly popular today). CiteSeerX (citeseerx.ist.psu.edu) is an example of an early digital library/academic search engine—founded in 1998 and focused on the computer and information science literature—that has over the years developed a number of social bookmarking, citation management and networking features.

From citation management to networking

The rapid uptake of social bookmarking and tagging has led to a parallel evolution from still popular, but traditional bibliographic citation management tools like EndNote or RefWorks toward more social citation management tools like Zotero (launched in 2006) and Mendeley (first release 2008) that are better integrated with web-based research and writing practices. Fenner's brief article (2010) comparing several citation management tools offers a simple

introduction from an end user perspective; Cohen (2008) outlines the history, novel approach, and significance of Zotero, while Puckett (2011) provides guidance how to effectively use it. Zaugg and others (2011) describe how Mendeley combines reference management with scholarly collaboration.

Citation management

Zotero is currently a popular tool supporting scholarly research, writing, citation and the personal organization or sharing of papers, reports, websites and blog posts, media and more. Its significance extends beyond its current functionality or user base. The principal achievement of the team that developed Zotero was to recognize and capitalize upon the interconnectedness of the social web to facilitate how scholars and students work. In keeping with the goal of the Center for History and New Media (chnm.gmu.edu) to combine scholarship and technology, the builders of Zotero have created a cloud-based researcher's tool to connect and integrate a disparate applications, services, repositories and content in a novel way. Zotero brings content and functionality from many sources together in the browser and enables not only its organization and storage, but also the easy synchronization and exchange of references and content.

Mendeley's public launch in 2008 was significant because the cloud-based, online version combines citation management with social networking and sophisticated new approaches to managing data. Mendeley extracts meaning from researchers' personal digital libraries (of citations and content) and patterns associated with the use of this content to reveal either additional articles of possible interest or colleagues with similar interests. Researchers can set up groups for collaboration, create professional profiles, and proactively connect and share content with other invited researchers. For some scholars, the impact of Mendeley has been

powerful; one of the digital library experts interviewed for this book said using it has increased his productivity and “changed much of my behavior.”

A blog post by political science professor Patrick Dunleavy (2012) captures some of the reasons for Mendeley’s popularity among its estimated two million users (Henning 2012). Arguably, as of this writing Mendeley may be among the largest repositories of academic papers: as of this writing its open catalog contains over 30 million papers, according to the Mendeley site (mendeley.com/how-we-help). It remains to be seen what the long term impact of Mendeley’s April 2013 acquisition by Elsevier will be; first reactions have varied (see for example Anderson 2013; Dobbs 2013; Howard 2013).

Regardless of what the future holds for the individual services, Zotero and Mendeley considerably broadened, smoothed, and quickened the pace on the highway toward scholarly social networks. Their existence is an instance of a broad shift to systems and services that *focus on scholars themselves*, in addition to the content they create. The success of Google Scholar—and especially the introduction of citation counts and links to cited articles, followed by the introduction of Google Scholar Citations in 2011 (Ortega and Aguillo 2012)—has further magnified the trend toward researcher-centered systems and online communities of scholars.

Scholarly collaboration on the social web

Researcher profiling systems

Scholarly researcher profiling systems increasingly use web-based tools to harvest information from disparate sources into expertise profiles for faculty, other researchers and even facilities like research labs. Some profiling systems are tied to national research assessment exercises (mentioned in the section on identifiers); others arose for other reasons.

VIVO

One early automated system dates to 2003, when a juxtaposition of disparate factors and opportunities at Cornell University led to the creation of VIVO (vivo.library.cornell.edu). VIVO was first implemented in 2004 as an online information service for providing an integrated view of the multidisciplinary, dispersed and disconnected life sciences community at Cornell (Devare et al. 2007; Corson-Rickert 2009, 67-79). The Cornell system now covers all disciplines.

A major early influence on VIVO was the internationally-supported Harmony Project, which brought fresh thinking to solving the complex interoperability problem of combining diverse metadata sets (Lagoze, Hunter and Brickley 2000). Rather than building a self-contained repository of digital content, VIVO's implementers built an index that functions as an overarching, unifying layer that highlights researchers and their interconnections: their affiliations, courses taught, grants, publications and more.

in 2009 an agency of the US National Institutes of Health (NIH) granted US\$12.2 million to build a new version of VIVO to support collaboration more broadly and enable scientists to network nationally (Gewin 2009; Brynko 2010; Krafft et al. 2010; Johnson, Buhler and Holmes 2010). At the time of this writing, instances of VIVO had been launched or are being implemented at over 50 institutions in North and Central America, Europe, Asia and Australia/New Zealand (wiki.duraspace.org/display/VIVO/VIVO+Main+Page).

Irrespective of its current position and prospects, VIVO's initial development was significant because it foreshadowed or exemplified certain social and technological shifts:

- The social web's shift of focus to people, collaboration and connections
- The recognition that an array of factors (in addition to refereed publications) determine a scholar's impact and influence in his or her field

- A shift from theory to practice for services based on semantic web technologies—specifically, modeling resources (both digital objects and real-world objects like people) as entities and relationships in ways that greatly facilitate their discovery, citation, re-use and re-aggregation on the web
- The possibility to move digital libraries from self-contained file systems (traditionally-conceived repositories) to service platforms for managing and providing access to diverse digital objects regardless of where they are stored
- Automated harvesting as the preferred means of collecting or pointing to content
- The critical importance of open, reusable and re-mixable data, not only for use within the site but also for disclosure to other sites

Difficulties and prospects for researcher profiling systems

Creating web-based services that profile or recommend experts is a difficult technical problem that has pushed the limits of digital library practice. It is a problem space that has attracted researchers in computer science. For example, in one often-cited article, Tang and others (2008) describe their work at Tsinghua University to automatically extract researcher profiles from the web, integrate publication data, and use the results to create an academic social network called ArnetMiner. Another example is the work reported by Fazel-Zarandi and others (2011). The authors, who note that the expert profiling process requires “reasoning about multiple complex networks from heterogeneous sources,” report a new framework for constructing expert profiles and recommender services. Their paper concludes with their planned next steps—to test the framework with researchers in particular domains, VIVO (vivoweb.org) and SciVal (scival.com).

Prospects for researcher profiling systems

It is early days for researcher profiling systems; from a functional perspective, there are alternatives; and it is difficult to predict how such systems will develop from here. Quite a few of them now exist. The Wikipedia entry that discusses them (“Comparison of Research Networking Tools and Research Profiling Systems” 2013) compiles information on over a dozen open-source, commercial, and institutionally managed tools or systems, among them VIVO and Digital Vita, which grew out of research at the University of Pittsburgh (Schleyer et al. 2008). Prominent commercial offerings in the Wikipedia article list include ResearcherID from Thomson Reuters (further discussed below) and SciVal from Elsevier.

Despite the achievements of researchers and implementers so far, it is not clear how much value scholars themselves place on researcher profiling systems. One of the digital library experts interviewed for this book worried that while he is aware of the enthusiasm of university administrators and librarians for one such system, he has “never heard a faculty member praise it.”

Marshak and Johnson (2010) conducted focus groups to look into faculty members’ perceptions of the value of researcher profiling systems. The focus groups turned up a number of perceived benefits of researcher profiling systems, but also a set of issues—which will be familiar to institutional repository managers—that could lead to low faculty engagement (see chapter 8). One (unsurprising) key finding was that faculty may be unwilling to keep their online profiles up to date. Another lesson for builders of academic profiling systems is that success may depend on a deep understanding and engagement with motivated, specific communities of scholars who have a stake in using the system for their own ends—and/or due to the requirements of national research assessment exercises, as discussed after the next section.

Scholarly social networks

Over the past three to five years, scholars have been adopting various tools associated with the social web. A number of studies have found that a majority of science and social science researchers report using non-academic social networks (Facebook), a growing number use services like LinkedIn, many are active on Twitter, and reading and commenting on academically-oriented blogs is fairly common (see for example Nentwick 2010; Giglia 2011; Chenu-Abente 2012; Gruzd and Goertzen 2013). At the same time, traditional scholarly publishing and conferences continue to remain much more important to researchers than social networks and media like blogs; the latest Ithaka studies of faculty preferences support this view, at least for US and UK researchers (Schonfeld, Housewright and Wulfson 2013; Housewright, Schonfeld and Wulfson 2013).

Existing networks

Nevertheless the success of social networks generally has attracted investment in a number of social platforms designed explicitly for scholars. Menendez, Angeli and Menestrina (2012, figure 4.1) list 19 social networks for researchers and their number of users as of October 2011. The aims of these freely available, public, network-based platforms are to support researchers' efforts to find information and research partners, keep up to date, contact or follow colleagues, form or work in groups, share or locate papers, and establish an online presence in their fields. Generally these services require scholars to create a profile, fill in a publication list (or "claim" their papers from a public database or databases), and identify their fields of interest.

Table 9.1 lays out summary information about the largest scholarly social networks as of early 2013: Academia.edu, ResearchGate, and Mendeley, which were all founded by web entrepreneurs who also have impressive credentials as scholars. Among these three, and as discussed previously, Mendeley is a kind of boundary object. It began as a citation management

tool (similar to Zotero) and has been evolving into a highly successful social network for researchers.

Table 9.1 Leading social networks specifically oriented to researchers 2013

Name	Founded /Location	CEO	Self-Reported Members/ Date	US Unique Visitors/ Month*	Funding
Academia.edu	2008 San Francisco	Richard Price	3.1 million (May 2013)	500,000+	Venture capital
ResearchGate.net	2008 Berlin	Ijad Madisch	2.8 million (May 2013)	350,000+	Venture capital
Mendeley.com	2008 London	Victor Henning (now VP Strategy, Elsevier)	2.0 million (November 2012)	125,000	Venture capital initially (acquired by Elsevier, April 2013)

*Source: compete.com; latest statistics available are for the month of March 2013 and represent US traffic only

Motivating factors for scholars

A question that is woven throughout this book also applies to these new social networks: what motivates or will motivate scholars to engage with them over the short and longer term? A study of social scientists' use of social sites suggest that the most-sought benefits are following other researchers' work, keeping up to date, discovering new ideas or publications, making new contacts, and promoting their own work (Gruzd and Goertzen 2013, 3338). The rapid growth of the three social networks discussed in this section suggests they are delivering these benefits and forming communities around their services.

At the same time, their claims of numbers of members are unconvincing. The UNESCO Institute for Statistics (2009) estimates there are around seven million researchers worldwide; could 3.1

million of them (44%) have signed up for Academia.edu? It seems unlikely. Results of the 2011 survey of Academia.edu users by Menendez and others (2012, 54), extrapolated to the 3.1 million members reported on the Academia.edu home page as of this writing, suggests that well over half a million members are undergraduates, alumni, retired faculty, or hold other types of positions; 1.25 million are graduate students; and about 1.3 million are faculty, post-docs, and independent researchers (still an impressive number).

Engagement, participation, incentives

Some studies also suggest that while membership numbers may be high in these three social networks, members' actual engagement and activity levels may be quite low. For example, Menendez and colleagues' analysis of Academia found a low level of user interaction and engagement with the service, with almost half of the members' never modifying their initial profiles (2012, 59). This is not a surprising finding; the challenge of getting users to actively participate in and contribute to online communities is not new. Despite their predominantly positive forecast for scholars' use of social media, Gruzdz and Goertzen conclude that "even though social media sites offer scholars a two-way form of communication and information exchange, scholars in the study tend to use social media in a one-directional mode" (2013, 3339).

Krichel and Zimmerman (2012), who have provided leadership for the RePEc open repository, discuss the critical importance of effective incentives to community engagement and participation (pointing in RePEc's case to its provision of an important author and institutional rankings service that is calculated from registration data). Government requirements related to national research assessments, research funding agency requirements, and repository deposit mandates are *external* incentives that influence scholars' choices and behaviors. In looking ahead to the prospects of scholarly social networks, the possibility of external incentives need to

be considered in combination with *internal* ones related to the life cycle of online communities (described in chapter 7). These include:

- How well scholarly social networks are able to align with the purposes and practices of multiple scholarly communities
- How successfully the builders can communicate and generate scholars' trust in their social network's value to scholarship
- How well aligned or embedded the social network is with other scholarly sites and requirements (for example, PubMed Central, arXiv, etc.; new requirements related to research assessments or researcher IDs, etc.)
- The perceived and actual benefits to visibility and prestige that they deliver
- How well they function (for example, how usable the sites are)
- How well they function as community centers across multiple disciplines
- What alternatives show up on the network over the next few years

Researcher identifiers

"In an ideal scholarly communication system there would be tools to browse, navigate, make recommendations and assess influence based on the complete graph of all actors (people, collaborations, institutions) and all communication artifacts (articles, comments, blog posts, usage data)." (Warner 2009)

The source of this remark by Simeon Warner, a research associate at Cornell with ties to the library, arXiv.org and the information science department, is a paper he presented at the 2009 International Conference on Open Repositories. His presentation, which describes the linked data approach of arXiv's author identifiers, also discusses other author identifier systems and the reasons they are important. One is that unambiguously identifying authors and associating

them accurately with their works may be the keystone of next-generation, global-scale collaboration for researchers on the social web.

Many have stressed the importance of developing author identifiers, often making the same points made decades ago for developing persistent identifiers for objects. For books, think of the benefits of an ISBN (International Standard Book Number); for articles, think of a DOI (Digital Object Identifier). Both exist to:

1. Unambiguously and persistently identify these entities
2. Make entities and objects related to them straightforward for machines and people to find and retrieve (see also the discussion of the Handle System in chapter 2)

Discoverability of researchers

Anyone who participates in the social web knows that it is can be far from straightforward to find and link to a single researcher, if the only way to search is that person's name. It is much harder to unambiguously identify people than other types of entities. Consider just a few aspects of the problem: many researchers have the same name (like Smith or Wang); the same individual's name can appear in various forms (e.g., with or without initials); and transliterations of non-Roman names yield variant spellings and word order. These types of problems are massively augmented in a global network environment in which millions of researchers' names are indexed.

In the context of scholarship's shift to the network, the challenges of accurately identifying particular scholars and attributing the results of scholarship to the right individual have become monumental. Julia Lane of the US National Science Foundation, writing for *Nature*, notes "on an international level, the issue of a unique researcher identification system is one that needs urgent attention" (2010, 488). Rotenberg and Kushmerick (2011), whose article provides

background on the author identification challenge, point to several factors driving the need for new solutions for scholarly name disambiguation, among them:

- The expansion of research output globally
- The rise of global collaborations and data sharing
- Increasing government or institutionally-based requirements to track research outcomes
- Large-scale growth in the number of researchers with Asian names, whose Roman transliterations are challenging to disambiguate (data from the UNESCO Institute for Statistics (2009) suggests that 38% of the world's estimated seven million researchers worldwide are in Asia, with half of these in China)

Library-based systems and VIAF

Libraries have a long tradition of community-wide systems (generally at the national level) for disambiguating names based on name authority control (the process of establishing the authoritative form of a name and linking variant forms as references, for the overall purposes of bringing the works related to that name together in the catalog, in turn facilitating search and retrieval). Niu (2013) describes name authority control in libraries, placing these practices in the context of recent initiatives inside and outside the library domain, and forecasting future directions for name authority control in libraries.

The most important recent library authority control-related developments in name identifiers include VIAF (the Virtual International Authority File; viaf.org), which leverages the investments in name authority control made by libraries all over the world. In 2009 VIAF became available as linked data, creating a new opportunity for the work of libraries to contribute to services that use semantic web approaches (Hickey 2010). A 2013 update on VIAF indicates it now brings together 27 national level authority files (Hickey 2013).

With respect to its contribution to author identification, VIAF links together different countries' authoritative forms of names, then assigns identifiers to the resulting clusters. These identifiers have subsequently been deployed as ISNIs (ISNIs are discussed below; see also Hickey 2011 and MacEwan, Angjeli and Gatenby 2013).

Will a single researcher identification system emerge?

Many individuals and types of organizations have a stake in unambiguously identifying authors and relating them to his or her scholarly contributions, collaborations or activities. Authors themselves have a stake, as do other researchers who are searching for related work, centers of expertise or potential collaborators. Libraries, funding agencies, research administrators, publishers and other aggregators of scholarly information have keen interest in attributing scholarly outcomes to the correct individuals.

Considering the social web, the deployment of a lightweight infrastructure to disambiguate names would make cross-domain searching, alerts, social bookmarking, citing and sharing content, social networking and researcher profiling services work more effectively and efficiently. Liu and colleagues (2012), who articulate a future-oriented perspective in their proposal for a research social network approach for analyzing local collaboration networks, would certainly be aided by reliable automated methods for unambiguously identifying authors and their works. There are excellent reasons to embrace and implement a unifying web-enabled system for uniquely and persistently identifying authors and reliably associating them with their work.

The use of author identifiers is still not common. The emergence of a single system for identifying authors is an unknown at the time of this writing. For now, many different author

identifier systems co-exist. Further, the most likely scenario is each author's having multiple identifiers. The rest of this section discusses a few of the better known alternatives that have been taken up in or across various communities. Martin Fenner's relatively brief overview of the author identification landscape (2011) is another source to consult.

Table 9.2 provides high-level summary data about the researcher identification services that are further discussed in the following subsections. The table lists only the better-known services and is not intended to be comprehensive.

Table 9.2 Selected researcher identification services

Name	Began	Type/Audience
VIAF (Virtual International Authority File)	2009	Library-based initiative, global in scope
ISNI (International Standard Name Identifier)	2010	Creators/contributors, producers, publishers ... ISO standard Global in scope
OpenID	2005	Authentication
DAI (Dutch Author Identifier)	2008	Dutch researchers
Lattes Platform	1999	Brazilian researchers
MIMAS Names Project	2007	UK researchers
ResearcherID	2007	Researchers worldwide
Scopus Author ID	2006	Researchers worldwide
arXiv Author ID	2005	Researchers in physics, math, computer science and several other disciplines
RePEc Author Service	1999	Researchers in economics and related sciences
AuthorClaim	2008	Researchers worldwide (based on RePEc's service)
ORCID (Open Contributor and Researcher ID)	2009	Researchers worldwide

General systems for identifying persons

ISNIs (International Standard Name Identifiers), which consist of 16 digits, are intended to disambiguate the identities by which individuals, characters and organizations are publicly known. ISNI is a certified global standard of the International Standards Organization (ISO) for "the unique identification of public identities across all fields of creative activity," including "the

millions of contributors to creative works and those active in their distribution, including writers, artists, creators, performers, researchers, producers, publishers, aggregators, and more” (isni.org/about). ISNIs, which function as a bridge to other name identification services and systems, form part of the family of identifiers that include the DOI and ISBN. The ISNI database contains or uses information drawn from a variety of data sources including VIAF, as noted previously (see also ISNI 2013). ISNI and ORCID (mentioned again later in this section) are complementary but distinct organizations. ORCID is focused on research and researchers and relies on researcher self-registration, while ISNI ingests, compiles and establishes identifiers from existing data sources across a number of domains.

OpenID (openid.net) is another standards-based, general-purpose system to unambiguously identify persons for the purpose of internet authentication. It is mentioned here for the purpose of clarification; OpenIDs do not have researcher profiling information associated with them; they are simply mechanisms for authentication. First developed by Brad Fitzpatrick (2005), OpenID enables users to sign on to the thousands of websites that accept OpenID authentication. The OpenID Foundation, an international non-profit membership organization established in 2007, is governed by sustaining, corporate and community members including among others Google, Facebook, Microsoft and PayPal (Thibeau 2011).

National systems and research evaluation

Some countries require the establishment of author profiles in support of national-level research evaluation and the showcasing of national research outputs (one example is Australia; see arc.gov.au). The following list briefly highlights the researcher identification systems associated directly with national research evaluation programs in four countries:

- Dutch Author Identifier (DAI), a unique identifier for each researcher in the Netherlands, linked to researcher profile pages (Jippes, Steinhoff and Dijk 2010)

- The government-supported, major source of information on Brazilian researchers, the Lattes Platform (lattes.cnpq.br), contains unique identifiers assigned by the CNPq system so that people with similar names are credited correctly. In 2010 Lattes contained data on about 1.6 million Brazilian researchers and students (Lane 2010; Hill 2011). It is the best example in the world of a mature and successful national system.
- National researcher identification systems are also well established in Norway (cristin.no/english) and Japan (http://rns.nii.ac.jp/html_us/help_en.html).

The UK agency JISC is investigating the possibilities for a national researcher identification system. The MIMAS Names Project (names.mimas.ac.uk), a collaboration of the University of Manchester and the British Library funded by JISC, explored and prototyped a service for UK repositories for uniquely identifying researchers. The project ran from 2007 to late 2011 (Cross et al. 2011) and delivered about 45 thousand identifiers for top UK researchers who were part of the national Research Assessment Exercise in 2008. Among other efforts, the project experimented with assigning ISNIs through machine matching, and it produced a plug-in for the EPrints repository software to help unambiguously associate researcher names with their work.

In a separate effort commissioned by JISC, Amanda Hill, the UK Names project manager, evaluated existing national researcher identifier systems (2011). She found that national systems tying the completion of researcher profiles to eligibility for research funding were the most successful at becoming integral parts of a nation's research infrastructure. Hill later presented on possibilities for future work (2012). She places the UK Names Project and the JISC Researcher Identifier evaluation in the context of other national systems and international initiatives such as ISNI and ORCID.

Publisher or repository-based systems

Leading scholarly publishers also provide identification systems. ResearcherID (researcherid.com) from Thomson Reuters is a freely available author identification service that functions at a global level (across national, publisher, institutional, research funding and disciplinary boundaries). While the service is associated with Thomson Reuters' Web of Science and Web of Knowledge, ResearcherID operates independently of them. Rotenberg and Kushmerick reported 127,000 ResearcherID profiles as of July 2011. An interactive map on the ResearcherID site suggests the service is currently most deployed in the United States, China, Brazil, Australia and across Europe. Elsevier also provides a researcher identification service called Scopus Author identifier, and ProQuest assigns researcher identifiers in connection with its large searchable database of faculty profiles, Scholar Universe.

A number of leading open repositories have author identification systems, with the most prominent being the arXiv author ID and the RePEc Author Service (authors.repec.org), which in mid-2012, registered 32,000 authors representing 85 to 90% of the top economists worldwide (Krichel and Zimmerman 2012). The code for the RePEc Author Service has been used to generate AuthorClaim (authorclaim.org), an author profiling and identification service with much broader scope across all disciplines.

Open Contributor and Researcher ID (ORCID)

CrossRef (crossref.org), an official DOI registration agency for scholarly publications, has been a strong advocate for a global, centralized author identification system. CrossRef convened an "AuthorID" meeting in early 2007 to discuss the possibilities for a registry and standard identification numbers for authors (Fenner 2009). Representatives from commercial and open access publishers, national libraries and library cooperatives and scholarly societies attended (crossref.org/CrossTech/2007/02/crossref_author_id_meeting.html), among them Jim Pringle of

Thomson Reuters, who spoke at the CrossRef meeting about ResearcherID (the code for ResearcherID was later licensed by ORCID).

The ORCID initiative started in 2009 following the first Name Identifier Summit convened by Nature Publishing Group and Thomson Reuters (CrossRef and KnowledgeSpeak 2009; Ratner 2012). ORCID, an open, nonproprietary and independent registry, went into operation in October 2012 (ORCID 2012). As of May 2013, there were 130,161 ORCID IDs (support.orcid.org/knowledgebase/articles/150557-number-of-orcid-ids). If ORCID succeeds in promoting the large scale uptake of identifiers by scholars and in making the many disparate researcher identification systems around the world more interoperable, it will be a massive force for advancing scholarly social networking in ways that will alter how scholars are identified, how they collaborate, how they are associated with their work, and how they are assessed.

The ORCID initiative has gained momentum across organizations, disciplines and countries in the past two years (Ratner 2011 and 2012; Haak et al. 2012). It is the most prominent current initiative, with a growing membership (orcid.org/about/community/members) providing financial support. The ORCID team has made a concerted effort to engage with the other organizations and initiatives discussed in this section (Ratner 2011, slides 28-29), with several represented on its board, as launch partners or on working groups.

ORCID is intended to bridge to and from other identification systems, supporting data exchange with ResearcherID, Scopus, RePEc, arXiv and others (Ratner 2012, slide 27). CrossRef, an ORCID launch partner, has continued to be a strong supporter of the initiative (crossref.org/10quarterly/quarterly.html), encouraging publishers to integrate ORCID IDs into their systems. ORCID IDs and ISNIs are interoperable 16-digit numbers, formatted to be compliant with the ISNI ISO standard. ORCID is using a block of numbers reserved for it by ISNI

(isni.org/isni_and_orcid). Infrastructure and organizational partnerships appear to be falling into place, but it remains to be seen if researchers themselves will be motivated to participate.

Conclusion

New notions of the web as a platform for participation have multiplied the possibilities for social digital libraries. When individuals who use social sites and tools approach digital libraries, they bring their social web expectations with them. The digital libraries that continue to operate from a traditional, collections-centered service model (that is, nearly all of them) are now faced with finding their place in the fast-moving, chaotic information space of the social web. There is nothing in the intrinsic nature of digital libraries blocking the way forward; most of the barriers arise from libraries' long-successful core values and assumptions. These values and assumptions sit somewhat uncomfortably beside the social web's core values: engagement and participation.

This chapter makes an attempt to visualize the key drivers and components of the evolution of digital libraries toward new roles on the social web. The chapter then turns to the consideration of the existing, emergent and possible impacts of the social web on researchers, scholarly tradition and scholarly practices. The social web is providing valuable complements to traditional methods for assessment, the identification of scholars and their works, and scholarly networking. A range of new social tools and services are now available to support research, keep up to date, learn about other scholars, track research outputs and more.

This chapter's analysis suggests that the social web offers many opportunities and benefits for scholars, but the transition may be slow or uneven due to motivational issues and scholarship's centuries-old, deeply entrenched traditions. Required participation in research assessments and other mandates aside, what will motivate scholars to engage with the new tools and social

networks over the short and longer term is an open question. Motivational factors have been a persistent barrier to the uptake of open repositories.

At the same time, open repositories are having a large impact on moving scholarship online in a way that makes it freely available to all. The publisher-provided scholarly literature has been online since 2008: a study by Cox and Cox (2008) reported that over 95% of science, technology and medicine and 85% of arts, humanities and social science journal titles were available in digital format. So, while almost all scholarly *literature* is online, so far the social web has not been a magnet for online *scholarly practice*.

The next and final chapter continues the consideration of digital libraries' evolution to the social web, this time in the context of new opportunities for digital library collections. That chapter concludes with some thoughts about what digital libraries can do for a society that is now so dominated by the social web.